Calculation of orthant probabilities by the holonomic gradient method

Tamio Koyama* and Akimichi Takemura†‡

January 21, 2013

Abstract

We apply the holonomic gradient method (HGM) introduced by [9] to the calculation of orthant probabilities of multivariate normal distribution. The holonomic gradient method applied to orthant probabilities is found to be a variant of Plackett's recurrence relation ([14]). However an implementation of the method yields recurrence relations more suitable for numerical computation than Plackett's recurrence relation. We derive some theoretical results on the holonomic system for the orthant probabilities. These results show that multivariate normal orthant probabilities possess some remarkable properties from the viewpoint of holonomic systems. Finally we show that numerical performance of our method is comparable or superior compared to existing methods.

1 Introduction

The holonomic gradient method (HGM) introduced by Nakayama et al. [9] is a new method of numerical calculation which utilizes algebraic properties of differential equations. This method has many applications in statistics. For example, an application to the evaluation of the exact distribution function of the largest root of a Wishart matrix is introduced in [5] and an application to the maximum likelihood estimation for the Fisher-Bingham distribution on the d-dimensional sphere is introduced in [7]. These applications greatly expand the scope of the field of algebraic statistics.

In this paper, we utilize the holonomic gradient method for an accurate evaluation of the orthant probability

$$\Phi(\Sigma, \mu) = \mathbf{P}(X_1 \geq 0, \dots, X_d \geq 0), \tag{1}$$

where the d-dimensional random vector $X = (X_1, \dots, X_d) \in \mathbf{R}^d$ is normally distributed with mean μ and covariance matrix Σ, i.e., $X \sim N(\mu, \Sigma)$. Since evaluation of the orthant

*Department of Mathematics, Kobe University
†Graduate School of Information Science and Technology, University of Tokyo
‡JST CREST

probability has important applications in statistical practice, there are many studies about it. Genz introduced a method to calculate the orthant probability utilizing the quasi Monte-Carlo method in [4]. Miwa, Hayter and Kuriki proposed a recursive integration algorithm to evaluate the orthant probability in [8].

When the mean vector μ is equal to zero, the orthant probability can be interpreted as the area of a $d-1$ dimensional spherical simplex (cf. [1]). In [16], Schläfli gave a classical differential recurrence formula for $\Phi(\Sigma, 0)$. Plackett generalized Schläfli's result and gave a recurrence formula for $\Phi(\Sigma, \mu)$ in [14]. He evaluated orthant probabilities by a recursive integration utilizing his formula. Gassmann implemented the Plackett's method in the case of higher dimensions in [3]. Their method is a recursive integration based on a differential recurrence formula, whereas the holonomic gradient method utilizes differential equations. Plackett's recurrence formula is not suitable for the holonomic gradient method and in this paper we give a new recurrence formula, which is more natural from the viewpoint of holonomic gradient method.

Let $y = \Sigma^{-1}\mu$ and $x = -\frac{1}{2}\Sigma^{-1}$. We denote the i-th element of y by y_i and the (i,j) element of x by x_{ij}. By this transformation of the parameters, the orthant probability in (1) can be written as

$$(-\pi)^{-d/2}\det(x)^{1/2}\exp\left(-\frac{1}{2}y^t x^{-1} y\right) g(x,y),$$

where

$$g(x,y) = \int_0^\infty \cdots \int_0^\infty \exp\left(\sum_{i,j=1}^d x_{ij}t_it_j + \sum_{i=1}^d y_it_i\right) dt \qquad (dt = dt_1 \dots dt_d). \tag{2}$$

In order to evaluate the orthant probability, it is enough to evaluate $g(x,y)$.

To apply the holonomic gradient method, we need an explicit form of a Pfaffian system ([7, Section 2]) associated with $g(x,y)$. The Pfaffian system can be obtained from a holonomic system for $g(x,y)$ (see [9]). For a given d, we can obtain a holonomic system for $g(x,y)$ by applying the algorithms introduced in [12] with computer algebra systems. However, for general d, these algorithms can not be applied and we need theoretical considerations. In this paper, we give a holonomic system for the function $g(x,y)$ and construct a Pfaffian system from the holonomic system.

Note that the integral $g(x,y)$ satisfies an incomplete A-hypergeometric system ([10]) when the integration domain is not the orthant but a polytope.

The organization of this paper is as follows. In Section 2, we describe a holonomic system for an integral

$$g(x,y) = \int_{\mathbf{R}^d} f(t)\exp\left(\sum_{i,j=1}^d x_{ij}t_it_j + \sum_{i=1}^d y_it_i\right) dt, \tag{3}$$

where $f(t)$ is a holonomic function or a holonomic distribution. In Section 3, we construct a holonomic system for the function in (2) by utilizing the result in Section 2. Then we construct a Pfaffian system from the holonomic system. Finally in Section 4 we describe numerical experiments of the holonomic gradient method.

2 Holonomic system associated with the expectation under multivariate normal distributions

In this section, we consider the holonomic ideal which annihilates the integral (3), which is the expectation under multivariate normal distributions (for the definition of the holonomic ideal, see [15]). We develop a general theory, where $f(t)$ in (3) is a smooth function or a distribution in the sense of Schwartz ([17]). This theory is a generalization of the result introduced in [6]. In Section 3 we will specialize $f(t)$ to be the indicator function of the positive orthant. Note that the results of this section can be applied to various problems of the multivariate normal distribution theory, other than the orthant probability.

We denote the ring of differential operators in x with polynomial coefficient by $D = \mathbf{C}\langle x_1, \ldots, x_n, \partial_{x_1}, \ldots, \partial_{x_n} \rangle$. The operator $\partial/\partial x_i$ is denoted by ∂_{x_i}. We frequently use the following rings:

$$\begin{aligned} D_{xyt} &:= \mathbf{C}\langle x_{ij}, y_k, t_k, \partial_{x_{ij}}, \partial_{y_k}, \partial_{t_k} : 1 \le i \le j \le d, 1 \le k \le d \rangle, \\ D_{xy} &:= \mathbf{C}\langle x_{ij}, y_k, \partial_{x_{ij}}, \partial_{y_k} : 1 \le i \le j \le d, 1 \le k \le d \rangle, \\ D_x &:= \mathbf{C}\langle x_{ij}, \partial_{x_{ij}} : 1 \le i \le j \le d \rangle, \\ D_t &:= \mathbf{C}\langle t_i, \partial_{t_i} : 1 \le i \le d \rangle. \end{aligned} \tag{4}$$

We use the following notation.

$$y = \Sigma^{-1}\mu, \quad x = -\frac{1}{2}\Sigma^{-1}, \quad h(x, y, t) = \sum_{i,j=1}^{d} x_{ij} t_i t_j + \sum_{i=1}^{d} y_i t_i. \tag{5}$$

2.1 The case of a smooth function

We first consider the case when $f(t)$ is a smooth function. In this case, the integral (3) converges if the matrix $-x$ is positive definite and $f(t)$ is of exponential growth.

To state the main theorem of this section, we show a general lemma. The notation f, g, x, y in the following lemma is generic and not related to (5).

Lemma 1. *Let D_x (resp. D_{xy}) be the ring of differential operators with polynomial coefficient $\mathbf{C}\langle x_i, \partial_{x_i} : 1 \le i \le n \rangle$ (resp. $\mathbf{C}\langle x_i, y_j, \partial_{x_i}, \partial_{y_j} : 1 \le i \le n, 1 \le j \le m \rangle$). Suppose that a holonomic ideal I annihilates a function $f(x, y)$ on $\mathbf{R}^n \times \mathbf{R}^m$ and the function $f(x, y)$ is rapidly decreasing with respect to the variable y for any x in an open set $O \subset \mathbf{R}^n$. Then the integration ideal of I with respect to the variable y annihilates*

$$g(x) := \int_{\mathbf{R}^m} f(x, y) dy \tag{6}$$

defined on the open set O.

Proof. Since the function $f(x, y)$ is rapidly decreasing, the integral of (6) converges. Let P be in the integration ideal $J = \left(I + \sum_{j=1}^{m} \partial_{y_j} D_{x,y}\right) \cap D_x$, then we have $P \bullet g = \int P \bullet f dy$

by the Lebesgue convergence theorem. Since the differential operator P can be written as

$$P = P_0 + \sum_{i=1}^{m} \partial_{y_i} P_i \in D_x \quad (P_0 \in I,\ P_i \in D_{xy}),$$

we have $\int P \bullet f dy = \sum \int \partial_{y_i} P_i \bullet f dy$. Since $P_i \bullet f$ is rapidly decreasing, we have $\int \partial_{y_i} P_i \bullet f dy = 0$. $\square$

We now go back to $g(x, y)$ in (3) and use the notation in (4) and (5). Consider a $\mathbf{C}$-algebra morphism φ from D_t to D_{xy} defined by

$$\varphi : D_t \to D_{xy} \qquad \left(t_i \mapsto \partial_{y_i},\ \partial_{t_i} \mapsto -y_i - 2 \sum_{k=1}^{d} x_{ik} \partial_{y_k} \right).$$

Here, we assume $x_{ij} = x_{ji}$, $\partial_{x_{ij}} = \partial_{x_{ji}}$. Since $[\varphi(\partial_{t_i}), \varphi(\partial_{t_j})] := \varphi(\partial_{t_i})\varphi(t_j) - \varphi(t_j)\varphi(\partial_{t_i}) = \delta_{ij}$, where δ_{ij} is Kronecker's delta, φ is well-defined as a morphism of $\mathbf{C}$-algebra.

Now we state the main theorem in this section.

Theorem 2. *Suppose a function f on $\mathbf{R}^d$ is smooth and of exponential growth. If differential operators $P_1, \ldots, P_s \in D_t$ annihilate f, then the following differential operators annihilate the integral $g(x, y)$ in (3).*

$$\varphi(P_k) \qquad (1 \le k \le s), \tag{7}$$

$$\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j} \qquad (1 \le i < j \le d), \tag{8}$$

$$\partial_{x_{ii}} - \partial_{y_i}^2 \qquad (1 \le i \le d). \tag{9}$$

Moreover, the differential operators (7), (8), (9) generate a holonomic ideal in D_{xy} if the differential operators $P_1, \ldots, P_s$ generate a holonomic ideal in D_t.

Proof. For a differential operator

$$P = \sum_{\alpha,\beta} c_{\alpha,\beta} t_1^{\alpha_1} \cdots t_d^{\alpha_d} \partial_{t_1}^{\beta_1} \cdots \partial_{t_d}^{\beta_d} \in D_t$$

and $p_i, q_i \in D_{xyt}$ $(i = 1, \ldots, d)$, we put

$$P(p_i; q_i) = \sum_{\alpha,\beta} c_{\alpha,\beta} p_1^{\alpha_1} \cdots p_d^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d}. \tag{10}$$

By the assumption, the differential operators

$$P_\ell \,(\ell = 1, \ldots, s), \quad \partial_{x_{ij}} \,(1 \le i \le j \le d), \quad \partial_{y_i} \,(1 \le i \le d)$$

annihilate f and generate a holonomic ideal in D_{xyt}. By the lemma introduced in [13, Section 3.3], the differential operators

$$P_\ell(t_i; \partial_{t_i} - \frac{\partial h}{\partial t_i}) \,(\ell = 1, \ldots, s), \quad \partial_{x_{ij}} - \frac{\partial h}{\partial x_{ij}} \,(1 \le i \le j \le d), \quad \partial_{y_i} - \frac{\partial h}{\partial y_i} \,(1 \le i \le d)$$

annihilate $\exp(h)f$ and generate holonomic ideal I in D_{xyt}. As we will show in Lemma 4 below, the differential operators (7), (8), (9) generate the integration ideal J of I with respect to the variables t_i $(i = 1, \ldots, d)$.

Since the function $\exp(h)\, f$ is rapidly decreasing, the integration ideal J annihilates the integral $g(x, y)$ by Lemma 1. Hence, the differential operators (7),(8), (9) annihilate $g(x, y)$.

Since the integration ideal of a holonomic ideal is also holonomic (see, e.g., [2, Chap 1]), the ideal J is holonomic when the ideal I is holonomic. □

Now, we show that the integration ideal J is generated by the differential operators (7), (8), (9) in the following two lemmas. We denote $P \equiv Q$ when $P - Q \in \sum_{i=1}^{d} D_{xyt}(\partial_{y_i} - t_i)$.

Lemma 3. *Let* $p_i = \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik} t_i$ *and* $q_i = \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik}\partial_{y_k}$. *For a differential operator* $P \in D_t$ *and a multi index* $\alpha \in \mathbf{N}_0^d = \{0, 1, 2 \ldots\}^d$, *the following equivalence relations hold.*

$$P(t_i; p_i) \equiv P(\partial_{y_i}; q_i), \tag{11}$$

$$t^{\alpha} P(\partial_{y_i}; q_i) \equiv \partial_y^{\alpha} P(\partial_{y_i}; q_i). \tag{12}$$

Here, we use the notation in (10).

Proof. By the straightforward calculation, we have the following relations for $1 \le i, j \le d$.

$$[p_i, q_j] = 0, \tag{13}$$

$$[p_i, p_j] = [q_i, q_j] = 0, \tag{14}$$

$$[q_i, t_j] = [q_i, \partial_{y_j}] = \delta_{ij}. \tag{15}$$

In order to prove (11), it is sufficient to show

$$t_1^{\alpha_1} \cdots t_d^{\alpha_d} p_1^{\beta_1} \cdots p_d^{\beta_d} \equiv \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d} \quad (\alpha, \beta \in \mathbf{N}_0^d). \tag{16}$$

By the induction on β_i, we have a relation

$$t_j q_i^{\beta_i} \equiv \partial_{y_j} q_i^{\beta_i} \quad (1 \le i, j \le d, \quad \beta_i \in \mathbf{N}_0). \tag{17}$$

When $\beta_i = 0$, the relation (17) holds clearly. Suppose that the relation (17) holds for β_i. Then we have

$$\begin{aligned} t_j q_i^{\beta_i+1} &= t_j q_i q_i^{\beta_i} = (q_i t_j - \delta_{ij}) q_i^{\beta_i} \\ &\equiv (q_i \partial_{y_j} - \delta_{ij}) q_i^{\beta_i} = \partial_{y_j} q_i q_i^{\beta_i} = \partial_{y_j} q_i^{\beta_i+1}. \end{aligned}$$

By induction, the relation (17) holds for any β_i.

By the relations (13) and (17), we have

$$p_1^{\beta_1} \cdots p_d^{\beta_d} \equiv q_1^{\beta_1} \cdots q_d^{\beta_d} \quad (\beta_i \in \mathbf{N}_0). \tag{18}$$

Now we prove the relation (16) by induction on the multi index $\alpha \in \mathbf{N}_0^d$. When $\alpha = 0$, the relation (16) holds because of (18). Suppose the relation (16) holds for α, then we have

$$\begin{aligned}
t_i t_1^{\alpha_1} \cdots t_d^{\alpha_d} p_1^{\beta_1} \cdots p_d^{\beta_d} &\equiv t_i \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d} \\
&= \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} (\prod_{j \neq i} q_j^{\beta_j}) t_i q_i^{\beta_i} \quad \text{(by (14) and (15))} \\
&\equiv \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} (\prod_{j \neq i} q_j^{\beta_j}) \partial_{y_i} q_i^{\beta_i} \quad \text{(by (17))} \\
&\equiv \partial_{y_i} \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d} q_1^{\beta_1} \cdots q_d^{\beta_d} \quad \text{(by (14) and (15))}.
\end{aligned}$$

By induction, the relation (11) holds for any α.

Finally, the relation (12) holds by $(14), (15)$ and (17). □

Lemma 4. *The integration ideal J is generated by the differential operators $(7), (8), (9)$.*

Proof. Note that the ideal I is generated by differential operators

$$\begin{aligned}
&P_\ell(t_i; \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik} t_i) \quad (\ell = 1, \ldots, s), \\
&\partial_{x_{ij}} - 2t_i t_j \quad (1 \leq i < j \leq d), \\
&\partial_{x_{ii}} - t_i^2, \ \partial_{y_i} - t_i \quad (1 \leq i \leq d).
\end{aligned}$$

By (11), the ideal I is generated by

$$P_\ell(\partial_{y_i}; \partial_{t_i} - y_i - 2\sum_{k=i}^{d} x_{ik} \partial_{y_k}) \quad (\ell = 1, \ldots, s), \tag{19}$$

$$\partial_{x_{ij}} - 2\partial_{y_i}\partial_{y_j} \quad (1 \leq i < j \leq d), \tag{20}$$

$$\partial_{x_{ii}} - \partial_{y_i}^2 \quad (1 \leq i \leq d), \tag{21}$$

$$\partial_{y_i} - t_i \quad (1 \leq i \leq d). \tag{22}$$

We denote by $\tilde{J}$ the left ideal generated by $(7), (8), (9)$. Clearly we have $\tilde{J} \subset J$. If P is a differential operator in J, then P can be written as

$$P = Q + \sum_{i=1}^{d} \partial_{t_i} R_i + \sum_{i=1}^{d} S_i \left(\partial_{y_i} - t_i\right) \quad (Q \in I, \ R_i, S_i \in D_{xyt}) \tag{23}$$

by the definition of integration ideal. The differential operator Q is written as a linear combination of the differential operators (19)–(21) with D_{xyt} coefficients. By the second term of the right-hand side of (23), we can assume without loss of generality that the variables ∂_{t_i} do not appear in these coefficients. By the relation (12) in Lemma 3, we can

assume that the coefficient in Q is an element in D_{xy}. Since any differential operator in (19)–(21) is an element of $\tilde{J} + \sum_{i=1}^{d} \partial_{t_i} \cdot D_{xyt}$, we can assume $Q \in J$.

Consider the equation

$$P - Q - \sum_{i=1}^{d} S_i(\partial_{y_i} - t_i) = \sum_{i=1}^{d} \partial_{t_i} R_i.$$

We can assume that the variables $\partial_{t_1}, \ldots, \partial_{t_d}$ do not appear in $S_1, \ldots, S_d$. For example, if $t_1\partial_{t_1}$ is a term of S_2 then we can replace S_2 and R_1 to $S_2 - t_1\partial_{t_1} - 1$ and $R_1 - t_1(\partial_{y_2} - t_2)$ respectively. The term $t_1\partial_{t_1}$ in S_2 is removed. In the same way, we can remove all terms which include ∂_{t_i}.

Expanding both sides and comparing the coefficients of ∂_{t_i}, we have

$$P - Q = \sum S_i(\partial_{y_i} - t_i).$$

The right-hand side of this equation is an element of the left ideal $J' := \sum_{i=1}^{d} D_{xyt} \cdot (\partial_{y_i} - t_i)$. Let the weight of t_i be 1 and that of other variables 0, and consider a term order $\prec$ with this weight. The set $\{t_i - \partial_{y_i} | 1 \leq i \leq d\}$ is a Gröbner basis of J' with the order, so that the initial term of $P - Q$ has to divide some t_i. Since $P - Q$ is in D_{xy}, we have $P - Q = 0$. Thus $P \in J$. $\square$

2.2 The case of a non-smooth function

Next we consider the case when $f(t)$ is not smooth. In this case we consider $f(t)$ as a distribution in the sense of Schwartz ([17]).

Let Ω be a domain defined by

$$\{(x, y) | -x \text{ is positive definite}\}.$$

For a tempered distribution f on $\mathbf{R}^d$, we can define a function on Ω as

$$g(x, y) = \langle f, \exp(h(t, y, x)) \rangle. \tag{24}$$

Since $\exp(h(t, y, x))$ is rapidly decreasing with respect to the variable t when $-x$ is positive definite, the right-hand side of (24) is finite.

A holonomic system for (24) is given as follows.

Theorem 5. *If differential operators $P_1, \ldots, P_s \in D_t$ annihilate a tempered distribution f on $\mathbf{R}^d$, then the differential operators* (7), (8), (9) *annihilate the function $g(x, y)$ in* (24). *Moreover, the differential operators* (7), (8), (9) *generate a holonomic ideal in D_{xy} if the differential operators $P_1, \ldots, P_s$ generate a holonomic ideal in D_t.*

Proof. We only need to prove that the differential operators (7), (8), (9) annihilate $g(x, y)$. Let $P_\ell = \sum c_{\alpha\beta} t^\alpha \partial_t^\beta$ and $P_\ell^* = \sum c_{\alpha\beta} t^\beta \partial_t^\alpha$. We have

$$P_\ell(\partial_{y_i}; -y_i - 2\sum_{k=i}^{d} x_{ik}\partial_{y_k}) \langle f, \exp(h(t, y, x)) \rangle$$

$$
\begin{aligned}
&= \langle f, P_\ell(\partial_{y_i}; -y_i - 2\sum_{k=i}^{d} x_{ik}\partial_{y_k}) \exp(h(t,y,x))\rangle \\
&= \langle f, P_\ell(\partial_{y_i}; -\partial_{t_i}) \exp(h(t,y,x))\rangle \\
&= \langle f, P_\ell^*(-\partial_{t_i}; \partial_{y_i}) \exp(h(t,y,x))\rangle \\
&= \langle f, P_\ell^*(-\partial_{t_i}; t_i) \exp(h(t,y,x))\rangle \\
&= \langle P_\ell(t_i; \partial_{t_i}) f, \exp(h(t,y,x))\rangle \\
&= 0.
\end{aligned}
$$

□

3 Holonomic system associated with the orthant probability

In this section we specialize $f(t)$ of the last section to the indicator function of the positive orthant and we will construct a Pfaffian system associated with the integral (2) for the orthant probability.

3.1 Generators of the holonomic ideal

At first, we obtain generators of a holonomic ideal which annihilates (2) by Theorem 5. Let E be the positive orthant in $\mathbf{R}^d$ defined by

$$\left\{ t = (t_1, \ldots, t_d) \in \mathbf{R}^d | t_i \geq 0, (i = 1, \ldots, d) \right\},$$

and $\mathbf{1}_E$ be the indicator function of E. A holonomic ideal which annihilates $\mathbf{1}_E$ is given as follows.

Lemma 6. *The indicator function $\mathbf{1}_E$ is annihilated by the following differential operators as a distribution.*

$$t_1\partial_{t_1}, \ldots, t_d\partial_{t_d} \tag{25}$$

The differential operators (25) *generate a holonomic ideal J in the ring D_t.*

Proof. At first, we show that the differential operators $t_i\partial_{t_i}$ annihilates the function $\mathbf{1}_E$. It suffices to prove for $i = 1$. Let $\varphi(t)$ be a rapidly decreasing function on $\mathbf{R}^d$. Then, we have

$$-\int_0^\infty \partial_{t_1}(t_1\varphi(t))dt_1 = 0$$

for any $t_j \in \mathbf{R}$, $2 \leq j \leq d$. Integrating both sides with respect to the variables $t_2, \ldots, t_d$, we have $\langle t_1\partial_{t_1}\mathbf{1}_E, \varphi\rangle = 0$. Therefore, the distribution $t_1\partial_{t_1}\mathbf{1}_E$ is equal to 0.

Next, we show that the left ideal J is holonomic. By the Buchberger's criterion, we can show that the set of the differential operators in (25) is a Gröbner basis of J with the

weight $w = (0, 1)$. The characteristic variety (see, e.g. [11], [15]) of J is $\mathrm{ch}(J) = \{(t, \xi) \in \mathbf{C}^{2d} | t_i \xi_i = 0,\ 1 \leq i \leq d\}$. The variety $\mathrm{ch}(J)$ can be decomposed as follows,

$$\bigcup_{J \subset \{1,\ldots,d\}} \{(t, \xi) | t_i = 0, \xi_j = 0, i \in J, j \notin J\}.$$

Since the Krull dimension of the each component is d, we have $\dim(\mathrm{ch}(J)) = d$ and the ideal J is holonomic. $\square$

The function $g(x, y)$ in (2) can be written as $g(x, y) = \langle \mathbf{1}_E(t), \exp(h(x, y, t)) \rangle$. This is a case of Theorem 5 in which the distribution f is $\mathbf{1}_E$. A holonomic ideal which annihilates $\mathbf{1}_E$ is given in Lemma 6. Hence we have the following theorem for $g(x, y)$ in (2).

Theorem 7. *Differential operators*

$$2 \sum_{k=1}^{d} x_{ik} \partial_{y_i} \partial_{y_k} + y_i \partial_{y_i} + 1 \quad (i = 1, \ldots, d, \quad x_{ij} = x_{ji}), \tag{26}$$

$$\partial_{x_{ij}} - 2 \partial_{y_i} \partial_{y_j} \quad (1 \leq i < j \leq d), \tag{27}$$

$$\partial_{x_{ii}} - \partial_{y_i}^2 \quad (1 \leq i \leq d) \tag{28}$$

annihilate the function $g(x, y)$ in (2), and generate a holonomic ideal I in the ring D_{xy}.

3.2 Differential recurrence formula

Next, we give a Pfaffian system associated with (2). In order to write the Pfaffian system, we define new notation. For $J \subset [d] = \{1, \ldots, d\}$, we put

$$h_J(x, y, x) = \sum_{i \in J} \sum_{j \in J} x_{ij} t_i t_j + \sum_{k \in J} y_k t_k,$$
$$g_J(x, y) = \int_0^\infty \cdots \int_0^\infty \exp(h_J(x, y, t))\, dt_J,$$

where $dt_J = \prod_{j \in J} dt_j$. When the set J is empty, we set $g_\emptyset = 1$. For example, the functions are written as follows when $d = 2$.

$$\begin{aligned}
g_{\{1,2\}}(x, y) &= \int_0^\infty \exp\left(t_1^2 x_{11} + 2 t_1 t_2 x_{12} + t_2^2 x_{22} + y_1 t_1 + y_2 t_2\right) dt_1 dt_2 \\
&= g(x, y), \\
g_{\{1\}}(x, y) &= \int_0^\infty \exp\left(t_1^2 x_{11} + y_1 t_1\right) dt_1, \\
g_{\{2\}}(x, y) &= \int_0^\infty \exp\left(t_2^2 x_{22} + y_2 t_2\right) dt_2, \\
g_\emptyset &= 1.
\end{aligned}$$

Let J be a subset of $[d]$ and put

$$\begin{aligned} Q_j &= -\left(y_j + 2\sum_{k=1}^{d} x_{jk}\partial_{y_k}\right) \quad (j=1,\ldots,d), && (29)\\ Q_{j,J} &= -\left(y_j + 2\sum_{k\in J} x_{jk}\partial_{y_k}\right) \quad (j\in J). && (30)\end{aligned}$$

Lemma 8. *The following equation holds.*

$$Q_j g_J = Q_{j,J} g_J = g_{J\setminus\{j\}} \quad (j \in J). \tag{31}$$

Proof. Since g_J is constant with respect to y_ℓ ($\ell \notin J$), we have $Q_j g_J = Q_{j,J} g_J$. We assume that $J = [d]$ without loss of generality. Applying Q_j to the integrand of $g_{[d]}$, we have

$$\begin{aligned} & -\left(y_j + 2\sum_{k=1}^{d} x_{jk}\partial_{y_k}\right) \exp\left(h(x,y,t)\right) \\ =& -\left(y_j + 2\sum_{k=1}^{d} x_{jk}t_k\right) \exp\left(h(x,y,t)\right) \\ =& -\partial_{t_j} \exp\left(h(x,y,t)\right). \end{aligned}$$

Integrating both sides of the equation from 0 to ∞ with respect to t_j, we have

$$Q_j \int_0^\infty \exp(h(x,y,t))dt_j = \exp\left(h_{[d]\setminus\{j\}}(x,y,t)\right).$$

Integrating both sides of the equation with respect to the remaining variables, we have the equation (31). $\square$

Remark. By Lemma 8, we have $Q_iQ_jg = g_{[d]\setminus\{i,j\}}$ for $i \neq j$ When the vector y is equal to zero, this equation can be written as

$$\left(2x_{ij} + 4\sum_{k=1}^{d} x_{ik}x_{jk}\partial_{x_{kk}} + 2\sum_{1\le k\neq \ell\le d} x_{ik}x_{j\ell}\partial_{x_{k\ell}}\right) g = g_{[d]\setminus\{i,j\}}.$$

By the transformation of parameters with $\Sigma = (\sigma_{k\ell}) = -\frac{1}{2}x^{-1}$, we have

$$\frac{\partial}{\partial \sigma^J_{k\ell}}\left(|\sigma|^{-1/2} g\right) = |\sigma|^{-1/2} g_{[d]\setminus\{i,j\}}.$$

It can be easily checked that this equation corresponds to the classical Schläfli's formula.

For $J = \{j_1,\ldots,j_s\} \subset [d]$, $(j_1 < j_2 < \cdots < j_s)$, we denote

$$\begin{aligned} x_J &= (x_{j_k j_\ell})_{1\le k,\ell\le s}, \quad y_J = (y_{j_1},\ldots,y_{j_s})^T, \\ \Sigma_J &= -\frac{1}{2}x_J^{-1} = (\sigma^J_{ij}), \quad \mu_J = \Sigma_J y_J = (\mu^J_{j_1},\ldots,\mu^J_{j_s}). \end{aligned}$$

The following differential recurrence formula holds.

Theorem 9. *For any $J \subset [d]$.*

$$\partial_{y_i} g_J \quad = \quad \begin{cases} \mu_i^J g_J + \sum_{j\in J} \sigma_{ij}^J g_{J\setminus\{j\}} & i \in J \\ 0 & i \notin J \end{cases} \tag{32}$$

$$\partial_{x_{ij}} g_J \quad = \quad \begin{cases} 2\partial_{y_i}\partial_{y_j} g_J & \{i,j\} \subset J,\, i<j \\ \partial_{y_i}^2 g_J & \{i\} \subset J,\, i=j \\ 0 & else. \end{cases} \tag{33}$$

Proof. For $i \notin J$, we have $\partial_{y_i} g_J = 0$ since the g_J is constant with respect to y_i. For $i \in J$ Lemma 8 implies

$$\begin{aligned} \mu_i^J g_J + \sum_{j\in J} \sigma_{ij}^J g_{J\setminus\{j\}} &= \mu_i^J g_J - \sum_{j\in J} \sigma_{ij}^J \left(y_j^J + 2\sum_{k\in J} x_{jk}^J \partial_{y_k} \right) g_J \\ &= \left(\mu_i^J - \sum_{j\in J} \sigma_{ij}^J y_j^J - 2\sum_{k,j\in J} \sigma_{ij}^J x_{jk}^J \partial_{y_k} \right) g_J. \end{aligned}$$

We have $\sum_{j\in J} \sigma_{ij}^J y_j^J = \mu_i^J$ by the relation $\mu^J = \Sigma^J y^J$ and we also have

$$-2\sum_{k,j\in J} \sigma_{ij}^J x_{jk}^J \partial_{y_k} = \sum_{k\in J} \delta_{ik}\partial_{y_k} = \partial_{y_i},$$

since $-2\Sigma^J x^J$ is the identity matrix of size $|J|$. Hence, the right-hand side of (32) equals $\partial_{y_i} g_J$.

For $\{i,j\} \not\subset J$, $\partial_{x_{ij}} g_J = 0$ since the g_J is constant with respect to x_{ij}. For $\{i,j\} \subset J$

$$\partial_{x_{i,j}} \exp\left(h_J(x,y,t)\right) = (2-\delta_{ij})\partial_{y_i}\partial_{y_j} \exp\left(h_J(x,y,t)\right).$$

Integrating both sides we have the relation (33). $\square$

By theorem 9, we have differential operators $\partial_{x_{ij}} - A_{ij}, \partial_{y_i} - A_i$ which annihilate the vector value function $G(x,y) = (g_J(x,y))_{j\subset[d]}$. Here, A_{ij} and A_i are $2^d \times 2^d$ matrices with rational function entries. In the next subsection, we prove that the system of these differential operators is a Pfaffian system in the meaning of [7, Section 2]. Note that the Pfaffian system have no singular point on

$$\{(x,y) | -x \text{ is positive definite }\}. \tag{34}$$

3.3 Pfaffian system and the holonomic rank

In this section, we give the holonomic rank of the ideal I generated by (26)–(28), and show that the differential recurrence formula (32) and (33) in the subsection 3.2 give a Pfaffian system associated with (2).

In this subsection, we denote the ring of differential operators in the variables x, y by R. The holonomic rank of I is the dimension of R/RI as a vector space over the field of rational functions $\mathbf{C}(x, y)$ (see, e.g., [9]).

At first, we give a lower bound of the holonomic rank. Since the holonomic rank equals to the dimension of holomorphic solutions of $I \bullet f = 0$ at generic points, we can obtain a lower bound of the holonomic rank by constructing linearly independent functions annihilated by I.

Lemma 10. *The holonomic rank of the ideal I in Theorem 7 is not less than 2^d, i.e.,* rank $I \geq 2^d$.

Proof. For a vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d) \in \{\pm 1\}^d$, let E_ε be an orthant

$$\left\{t = (t_1, \ldots, t_d) \in \mathbf{R}^d | \varepsilon_i t_i > 0 \ (i = 1, \ldots, d)\right\}.$$

It is enough to show that the following 2^d functions are linearly independent and annihilated by I;

$$g_\varepsilon(x, y) = \int_{E_\varepsilon} \exp\left(h(x, y, t)\right) dt.$$

By an analogous way as in the proof of Lemma 6, we can show that the indicator function of E_ε is annihilated by the differential operators in (25) for any $\varepsilon \in \{\pm 1\}^d$. Analogously to the proof of Theorem 5, we can show that the differential operators (26)–(28) annihilate $g_\varepsilon(x, y)$.

Let c_ε be a real number for $\varepsilon \in \{\pm 1\}^d$, and suppose $\sum_\varepsilon c_\varepsilon g_\varepsilon = 0$. Multiplying both sides of the equation by $(2\pi)^{-d/2} (\det \Sigma) n^{-1/2} \exp(-\frac{1}{2} \mu^t \Sigma^{-1} \mu)$, we have

$$\sum_{\varepsilon \in \{\pm 1\}^d} c_\varepsilon \mathbf{P}(E_\varepsilon | \mu, \Sigma) = 0.$$

Here, $\mathbf{P}(E_\varepsilon | \mu, \Sigma)$ is the probability of the event E_ε under the multivariate normal distribution $N(\mu, \Sigma)$. Substituting $\mu = t\varepsilon, (\varepsilon \in \{\pm 1\}^d)$ and taking a limit $t \to +\infty$, we have

$$c_\varepsilon = 0$$

Hence, the functions $g_\varepsilon(x, y)$ are linearly independent. □

In order to obtain an upper bound of the holonomic rank of I, we construct bases of R/RI as a linear space over $\mathbf{C}(x, y)$. The bases correspond to the functions g_J.

For $J \subset [d]$, we put a differential operator

$$P_J = \prod_{j' \in [d] \setminus J} Q_{j'}, \tag{35}$$

where Q_j is the differential operator in (29). Note that the differential operators Q_j commute with each other. By Lemma 8, we have

$$P_J g = g_J. \tag{36}$$

Equation (36) means that the differential operator P_J corresponds to the function g_J. For example, When $d=2$ and $J=\emptyset$, the equation (36) is written as follows.

$$(y_1+2x_{11}\partial_{y_1}+2x_{12}\partial_{y_2})\,(y_2+2x_{21}\partial_{y_1}+2x_{22}\partial_{y_2})\,g_{\{1,2\}}=1.$$

Since the differential operator Q_j commutes with $\partial_{x_{ij}}-2\partial_{y_i}\partial_{y_j}$ $(1\le i<j\le d)$ and $\partial_{x_{ii}}-\partial_{y_i}^2$ $(1\le i\le d)$, we have the following lemma.

Lemma 11. *The following formulas hold in R/RI.*

$$\partial_{x_{ij}}P_J=2\partial_{y_i}\partial_{y_j}P_J \quad (1\le i<j\le d, J\subset[d])\,, \tag{37}$$
$$\partial_{x_{ii}}P_J=\partial_{y_i}^2P_J \quad (1\le i\le d, J\subset[d])\,. \tag{38}$$

Proof. For $1\le i<j\le d$ and $1\le k\le d$, we have

$$\partial_{x_{ij}}Q_k=-\partial_{x_{ij}}\left(y_k+2\sum_{\ell=1}^d x_{k\ell}\partial_{y_\ell}\right)=Q_k\partial_{x_{ij}}-2\delta_{ik}\partial_{y_j}-2\delta_{jk}\partial_{y_i},$$
$$2\partial_{y_i}\partial_{y_j}Q_k=-2\partial_{y_i}\partial_{y_j}\left(y_k+2\sum_{\ell=1}^d x_{k\ell}\partial_{y_\ell}\right)=Q_k2\partial_{y_i}\partial_{y_j}-2\delta_{ik}\partial_{y_j}-2\delta_{jk}\partial_{y_i}.$$

For $1\le i\le d$ and $1\le k\le d$, we have

$$\partial_{x_{ii}}Q_k=-\partial_{x_{ii}}\left(y_k+2\sum_{\ell=1}^d x_{k\ell}\partial_{y_\ell}\right)=Q_k\partial_{x_{ii}}-2\delta_{ik}\partial_{y_i},$$
$$\partial_{y_i}^2Q_k=-\partial_{y_i}^2\left(y_k+2\sum_{\ell=1}^d x_{k\ell}\partial_{y_\ell}\right)=Q_k\partial_{y_i}^2-2\delta_{ik}\partial_{y_i}.$$

Therefore, the differential operator Q_j commutes with $\partial_{x_{ij}}-2\partial_{y_i}\partial_{y_j}$ $(1\le i<j\le d)$ and $\partial_{x_{ii}}-\partial_{y_i}^2$ $(1\le i\le d)$. Then we have

$$(\partial_{x_{ij}}-2\partial_{y_i}\partial_{y_j})P_J=P_J(\partial_{x_{ij}}-2\partial_{y_i}\partial_{y_j})=0$$

in R/RI. Similarly we have $(\partial_{x_{ii}}-\partial_{y_i}^2)P_J=0$. $\square$

The following lemma corresponds to Lemma 8.

Lemma 12. *With the same notations as in Lemma* 8*,*

$$Q_jP_J=Q_{j,J}P_J=P_{J\setminus\{j\}} \quad (j\in J,\ J\subset[d]). \tag{39}$$

holds in R/RI.

Proof. By definition of P_J, it is clear that $Q_j P_J = P_{J\setminus\{j\}}$. Since $\partial_{y_\ell} Q_\ell \in I$, $\partial_\ell P_J$ is in I if $\ell \notin J$. Hence we have

$$Q_j P_J = -\left(y_j + 2\sum_{k=1}^{d} x_{jk}\partial_{y_k}\right)\prod_{\ell\in[d]\setminus J} Q_\ell = Q_{j,J} P_J$$

in R/RI. Note that Q_j and ∂_{y_k} commute if $j \neq k$. □

Theorem 13. *The following relations hold in R/RI for any $J \subset [d]$.*

$$\partial_{y_i} P_J \equiv \begin{cases} \mu_i^J P_J + \sum_{j\in J} \sigma_{ij}^J P_{J\setminus\{j\}} & i \in J \\ 0 & i \notin J, \end{cases} \tag{40}$$

$$\partial_{x_{ij}} P_J \equiv \begin{cases} 2\partial_{y_i}\partial_{y_j} P_J & \{i,j\} \subset J,\ i<j \\ \partial_{y_i}^2 P_J & \{i\} \subset J,\ i=j \\ 0 & else. \end{cases} \tag{41}$$

Proof. For $i \in J$, we can show the equation (40) by Lemma 12 and an analogous calculation as in the proof of (32). For $i \notin J$, the equation (40) is shown as in the proof of Lemma 12. By Lemma 11, we have the equation (41). □

Corollary 14. *The set of the differential operators P_J ($J \subset [d]$) in* (35) *spans the quotient space R/RI as a vector space over* $\mathbf{C}(x,y)$.

This corollary together with Lemma 10 establishes the following theorem.

Theorem 15.

$$\mathrm{rank}\ I = 2^d$$

4 Numerical experiments

In this section we present numerical experiments of our holonomic gradient method for orthant probabilities. Our experiments show that the holonomic gradient method is very accurate and fast compared to existing methods.

By theorem 9, we have an explicit formula for $\frac{\partial}{\partial t}G(x(t),y(t))$ when $x(t)$ and $y(t)$ are smooth functions. In order to evaluate $G(x,y)$ at (x_1,y_1), we put

$$x(t) = (1-t)x_0 + tx_1,\ \ y(t) = ty_1 \quad 0 \le t \le 1. \tag{42}$$

Here, x_0 is the diagonal matrix whose (i,i)-entry equals to that of x_1. The initial value can be written as

$$g_J(x(0),y(0)) = \prod_{j\in J}\left(-\frac{\pi}{4}x_{jj}(0)\right)^{\frac{1}{2}}. \tag{43}$$

Note that $\frac{\partial}{\partial t}G(x(t),y(t))$ does not have singular points on $[0,1]$ since the Pfaffian system for $G(x,y)$ does not have singular point on (34).

Table 1: Errors

No.	dim	error
1	2	1.760124e-08
2	3	5.473549e-08
3	4	3.373671e-08
4	5	2.265284e-09
5	6	1.120033e-08
6	7	7.330036e-09
7	8	8.705609e-09
8	9	2.288549e-09
9	10	5.024879e-10

The accuracy of the holonomic gradient method can be checked by looking at the summation $\sum_{\varepsilon\in\{\pm1\}^d}\mathbf{P}(X\in E_\varepsilon)=1$. Table 1 shows errors $|1-\sum_{\varepsilon\in\{\pm1\}^d}\mathbf{P}(X\in E_\varepsilon)|$ for sample data.

For a correlation matrix $R=\{\rho_{ij}\}$ with $\rho_{ij}=\rho, i\neq j$, the orthant probability can be written as a one-dimensional integral. Table 2 shows the values of the orthant probability calculated by the one-dimensional integral and HGM for certain values of ρ. The differences of the result of HGM and the one-dimensional integral are less than 10^{-06}.

Table 2: Comparing of one-dimensional integral

rho	Dunnett	HGM	error
0.0	0.00097656249807343551	0.000976562500000	1.926564e-12
0.1	0.0065864743711607976	0.006586475124405	7.532442e-10
0.25	0.026603192349344017	0.026603192572268	2.229240e-10
0.5	0.090909089922689604	0.090909086078297	3.844393e-09

Table 3 shows averages of computational times of holonomic gradient method and Miwa's method for 100 sample data. The mean vectors of the sample data are all zero, and the covariance matrices are randomly generated. Table 3 shows that our holonomic gradient method evaluates orthant probabilities faster than Miwa's method, as the dimension becomes larger.

However, it should be noted that the holonomic gradient method can be very slow when the mean vector is far away from zero. When the mean vector is far away from zero and some eigenvalue of the covariance matrix is very small, the value of the integral (2) becomes very large since the parameter y becomes large. In such a case, the Runge-Kutta method takes a long time.

Table 3: Averages of computational times

dim	Miwa	HGM	dim	Miwa	HGM
5	0.002	0.016	9	6.078	1.050
6	0.011	0.056	10	60.171	2.371
7	0.080	0.154	11	671.370	5.411
8	0.664	0.390	12	-	13.48

References

[1] K. Aomoto, *Analytic structure of Schläfli function*, Nagoya Math.J., **68**,1–16, 1977.

[2] J. E. Björk, *Rings of Differential Operators*, North-Holland, New York, 1979.

[3] H. I. Gassmann, *Multivariate normal probabilities: implementing an old idea of Plackett's*, Journal of Computational and Graphical Statistics, **12**, No.3, 731–752, 2003.

[4] A. Genz, *Numerical computation of multivariate normal probabilities*, Journal of Computational and Graphical Statistics, **1**, 141–150, 1992.

[5] H. Hashiguchi, Y. Numata, N. Takayama, and A. Takemura, *Holonomic gradient method for the distribution function of the largest root of a Wishart matrix*, arXiv:1201.0472, 2012.

[6] T. Koyama, *A holonomic ideal annihilating the Fisher-Bingham integral*, arXiv:1104.1411, 2011.

[7] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama, *Holonomic gradient descent for the Fisher-Bingham distribution on the d-dimensional sphere*, arXiv:1201.3239, 2012.

[8] A. Miwa, J. Hayter, and S. Kuriki, *The evaluation of general non-centered orthant probabilities*, Journal of the Royal Statistical Society, Ser. B, **65**, 223–234, 2003.

[9] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, and A. Takemura, *Holonomic gradient descent and its application to the Fisher-Bingham integral*, Advances in Applied Mathematics, **47**, 639–658, 2011.

[10] K. Nishiyama and N. Takayama, *Incomplete $\mathcal{A}$-hypergeometric systems*, T. Hibi editor, Harmony of Gröbner Bases and the Modern Industrial Society, World Scientific, Singapore, 193–212, 2012.

[11] T. Oaku, *Computation of the characteristic variety and the singular locus of a system of differential equations with polynomial coefficients*, Japan Journal of Industrial and Applied Mathematics, **11**, 485–497, 1994.

[12] T. Oaku, *Algorithms for integrals of holonomic functions over domains defined by polynomial inequalities*, Journal of Symbolic Computation,**50**, 2013.

[13] T. Oaku, Y. Shiraki, and N. Takayama, *Algorithms for D-modules and numerical analysis*, Z. M. Li, W. Sit (editors), Computer Mathematics, World Scientific, Singapore, 23–39, 2003.

[14] R. L. Plackett, *A reduction formula for normal multivariate integrals*, Biometrika, **41**, No.3/4, 351–360, 1954.

[15] M. Saito, B. Sturmfels, and N. Takayama, *Gröbner Deformations of Hypergeometric Differential Equations*, Springer, 2000.

[16] L. Schläfli, *On the multiple integral $\int^n dxdy \ldots dz$ whose limits are $p_1 = a_1x + b_1y + \cdots + h_1z > 0, \quad p_2 > 0, \ldots, p_n > 0$ and $x^2 + y^2 + \cdots + z^2 < 1$*, Quart. J. Pure Appl. Math. **2**, 269–301; **3**, 54–68, 97–108, 1858.

[17] L. Schwartz, *Théorie des distributions*, Hermann, Paris, 1966.